ITCR Annual Meeting, June 13-14, 2016, Broad Institute, Cambridge MA

Meeting Notes with Focus on Identifying Challenges, Needs and Collaboration Opportunities

Twitter #nciitcr

Day 1

Key Note

Corrie Painter - Empowering Patients to Drive Biomedical Research

- Angiosarcoma survivor and scientist
- Found more information about angiosarcoma on Facebook than in PubMed or by Google search
- More impact as an advocate - ability to influence research directions
- Approach to patient-driven genomics - IRB approved research protocol
    - Online consent form
    - Medical records access
    - Tissue samples obtained
    - Data will be deposited into Genomic Data Commons
- Metastatic Breast Cancer Project.org
- Close patient engagement in web site graphic design and language
- Launched 10/2015; participation in all 50 states. Patient-driven outreach via social media is most effective for enrollment
- Patient interaction is a priority - regular updates via website and email; send educational information and notifications of new groups being studied, receive feedback about questions and suggestions.
- "Saliva selfie" became a badge of pride for participation. Patients were invited to the Broad to meet with the cancer program researchers
- Patient reported data: 16 questions answered in an average of 6 minutes. 95% submit the full survey. >2200 patients completed. >1200 consented; >500 saliva received
- Focus on patients with exceptional responses
- 35% of participants are under the age of 40 - expected due to online approach
- Major goal for 2016 is to increase racial and ethnic diversity
- Broad exploring if other projects (e.g. rare cancers) could be modeled after the MBC project:
- Angiosarcoma project being developed now, modeled after the MBC project

Genomics Session I

Martin Morgan: Cancer Genomics: Integrative and Scalable Solutions in R / Bioconductor

Three key aspects to the ITCR project:

- 1) New object for facilitating work with multiple assays on similar individuals: MultiAssayExperiment

- 2) Concept of Hubs for very easy access to 'Annotation' (consortium-level information to help inform individual analyses) and 'Experiment' (curated summaries of exemplary data) resources. Building a package for connecting to the Genomic Data Commons.
- 3) Working with large data: Representations of data so that work can be more interactive until the actual analysis takes place; simplified transition from local computation to cluster or cloud compute computations (single command).

Jingchun Zhu - UCSC Xena - http://xena.ucsc.edu

Two components:

1) Federated data hubs. All information but the sequence level data (not BAMs, FASTQs)
    a) Private data hubs (e.g., laptops)
    b) Public data hubs. Hosted by UCSC, large datasets (e.g. TCGA, ICGC, TARGET)
2) Xena Browser: links data across public and private hubs

Collaborating projects:

1) GA4GH BD2K RNASeq Recompute
   - Performed by TOIL
2) PCAWG Pan Cancer Analysis of Whole Genome datasets
3) Interpretation of N=1 case, use of Xena composite cohort

Jill Mesirov: GenePattern Notebooks for Cancer

- Over 200 GenePattern modules, analytical pipelines, connects to many software tools including six ITCR tools.
- Gene Pattern Notebooks available since February 2016
- GenePattern now licensed under BSD3
- Notebook: Focus on non-programmers, seamless integration of GenePattern with Jupyter notebook.
- Jupyter == iPython
- Use case: prediction of medulloblastoma outcome
- Notebook repository available, no installation required.
- Discover cancer models.

Trey Ideker: NDEx, the Network Data Exchange http://www.ndexbio.org

- V1.3 currently focused on community, content and connections to cancer network analysis
- Types of network analysis:
  1) gene- or variant centered
  2) patient-centered (N of 1; patient-similarity networks)
- Network-based stratification: places patients into clinically informative subtypes
- CX: new data exchange format for networks
- CyNDEx: transfer data between NDEx and Cytoscape
- Journal Links: Networks are published as data. The public NDEx site is used for network submission and review. Collaboration with Elsevier, issuing DOIs for networks.
- Massive guided network curation in collaboration with Peter Sorger (DARPA).

- Forming a NDEx Cancer Advisory Board
- Demo is focused on advanced search capabilities

Imaging Session I

David Gutman & Lee Cooper: Cancer Digital Slide Archive - Infrastructure and algorithms for management, analysis and integration of digital pathology data
http://cancer.digitalslidearchive.net/

- Kitware as developer
- Challenge: dealing with PHI. Need to keep this information siloed
- HIstomicsTK for Whole Slide Image Analysis
- GitHub auto-build with upload to Docker Hub
- Looking towards linking markups to actual ontologies
- HistomicsTK demo: Generation of heatmaps for cell types in a sample. Interactive learning method to refine classifiers (to come online in the next year)

Jayashree Kalpathy-Cramer: Tools for Optimized Imaging Biomarkers for Cancer Research and Discovery https://cbibop.github.io/

- Resources to coordinate and conduct challenges including metrology tools and tools
- C-BIBOP: Cloud-based imaging biomarker optimization platform
- MICCAI brain tumor segmentation challenge - Run since 2012; organized by academics and NCI
- New paradigm: moving code to the data; approaches such as Docker to facilitate
- Shiny/R integration for statistical analysis
- Cloud-based imaging workflows
- Looking into using Intel personalized care platform

Andrey Fedorov: Quantitative Image Informatics for Cancer Research (QIICR) (pronounced "quicker") http://qiicr.org/

- Only 1 FDA-qualified imaging biomarker! ~100 routinely used in the clinic
- NCI QIN: sharing mostly taking place among members of this community
- Goal of QIICR: Standards-based sharing of analysis results and tool sharing
- All tools are "no strings attached" open source. Tools for developers and non-developers
- DICOM for imaging and non-imaging data
- Example: Tumor metabolic activity as a biomarker. Communication of this information includes: Patient weight and injection dose for normalization; identification of Region of Interest (ROI); feature extraction. QIICR tools provides tools for all of these steps.
- Recent paper in PeerJ http://peerj.com/articles/2057/
- Created tools to simplify interaction with DICOM
- Development of 3D SlicerPathology extension for digital pathology image annotation (PI: Saltz)
- Community Building: NA-MIC Project Week "Hack-a-Thon". @MIT in January
- Video available here: https://youtu.be/wK2TGyVQjzs and also from the project main page: http://qiicr.org

Christos Davatzikos: Cancer Imaging Phenomics Software Suite

- Integrates a number of different algorithms and projects
    1) family of advanced imaging computing algorithms
    2) leverage collaborations with clinical teams
- Main Elements of CaPTK
    - Image analysis algorithms
    - Precision diagnostics using machine learning
    - Personalized treatment planning
- Examples: Radiomic breast cancer phenotypes; brain connectivity
- LIBRA: Fully automated breast density estimation

Notes during TOW meeting

- Should we develop an ITCR "playlist" on YouTube?
- Surveys: Use of social media for outreach activities. Andrey will make this available to new grantees.
- Discussion of Journal Special Issue
    - Guidelines for authors: target audience is cancer researchers
    - Coordinated press releases - tricky. TCGA model: Lead press release from the NCI, participating institutes to follow.
    - CI4CC-get their input on what would be impactful?
    - Genomic Data Commons overview paper?
    - Need to be able to point to technical supporting information
        - Define what questions each resource needs to include in the supplemental information.
    - Are there any access issues with data resources?
    - Address with potential issue of NCI endorsement.
    - Should there be a NCI context? - Broader NCI informatics strategies including service to NCI programs.
- For next year:
    - Invite a guest speaker to talk about outreach, even outside of bioinformatics
    - How do we measure success? -
    - How do we track usage? Genome Space example. Aggregate statistics. IGV: tracks each launch of the tools (home call)
    - "Amazon" of tools? Would this be valuable? "If Google doesn't know about your tool, it doesn't exist."
    - Domain-specific communities, see BioDocker as an example

Lightning Talks

- Xiao Xinshu: informatics tools for single nucleotide analysis of RNA-Seq data RASER, GIREMI, eGRAS, iGRAS ExpressGV
- Chakra Chennubhotla Srinivas: tumor heterogeneity in multiplexed fluorescence images.
- Terry Meehan: PDX integrator, dissemination of data from patient derived xenografts
- Daniel Marcus: integrative imaging informatics for cancer research

- Hua Xu: advancing cancer pharmacological research through EHRs and informatics
- Timothy Griffin: Multi-omics hub for cancer research (Galaxy)

Genomics Session II

Brian Haas: Trinity, tools for transcriptome assembly
https://github.com/trinityrnaseq/trinityrnaseq/wiki

- De novo transcriptome assembly. 1.5K unique users per month.
- OS development from Trinity community
- Google Group, Twitter, workshops, galaxy
- Toolkit: e.g. mutations, fusion transcripts, lincRNAs, viruses & microbes, single cell tumor heterogeneity, etc.
- SLNCKY: LincRNA identification from reconstructed transcripts (https://slncky.github.io)
- InferCNV: ID large scale CNV from RNASeq data
- Integration of Trinity CTAT into WDL workflows to run on FireCloud - to run on 1000's of samples. Single cell tools are a separate package.
- Collaboration with the Broad Inst. clinical sequencing program
- Question: what are the collaboration opportunities with the Krasnitz project?

Alexander Krasnitz: Computational framework for single cell genomics of tumors

- Profile very large number of single cells from individual tumors, only about 5% of the genome is sequenced, plan to increase the amount of DNA that is sequenced/nuclei.
- Viewer created as a Docker image
- Example of clinical utility: single-cell genomic profiling of prostate cancer. Current approach has shortcomings. 65% that any 2 pathologists disagree; differing scores on core vs post-RP biopsies; some low-scoring cases may be aggressive due to sub-clonal cell populations that go undetected.
- Used single cell analysis + viewer to reveal highly complex cellular lineages.
- Hope to go live with the tool later this year. Current adopters @MSKCC and @NYGC
- Working to Dockerize the remainder of the pipeline
- Early detection: detect clonal populations among cells in circulation, following depletion of nucleated blood cells
- >Question: are there collaboration opportunities with the new Srinivas project?

Xiaole Shirley Liu: Developing informatics technologies to model cancer gene regulation

- Cistrome analysis pipeline http://www.cistrome.org/Cistrome/Cistrome_Project.html
- Factors influencing cancer-specific transcription factor targets. E.g., Regulatory potential influenced by distance from the TSS (exponential decay)
- MARGE uses logistic regression to explain differential expression gene set of interest
- Website: TCGA Cancer Enhancer Prediction
- Video at https://www.youtube.com/watch?v=Pel0Crb6ZLY

Tech WG Demonstration Projects

Tahsin Kurc: Containers and APIs WG Demonstration Project

- Can we take advantage of the NCI Cloud projects?
- Part 1: Dockerization of image analysis codes, on demand analysis of images and feature extraction.
    - Supports reproducibility. Container holds all software dependencies, including proper versions. Using the SB Genomics Cloud Pilot as a platform for this work
- Part 2: Dissemination of results
    - Features in the BigQuery table, working with the ISB Google Cloud
    - Clustered Heatmap can work on top of of BigQueryr

- Question: Would it be possible to extract the Docker container and run outside commercial grid? Yes.
- Comments: There are some security issues with Docker. Cost is an issue with Cloud Computing, NCI is looking into this. Cost/time: should urgent cloud computing jobs cost more, discount if the job can wait?

Clinical Session

Guagian Jiang: caCDE-QA, a quality assurance platform for cancer study common data elements

- Harmonization of data elements between different studies
- caDSR based on ISO metadata standards, data dictionary using caDSR.
- Amis: develop OWL-based QA tools, apply QA tools, deploy QA tools
- HL7 Fast Healthcare Interoperability Resources (FHIR)
- Shape expression language ShEx
- D2Refine: a metadata harmonization and validation frame work.
- Two posters at the meeting

Rebecca Jacobson: Advanced development of TIES http://ties.dbmi.pitt.edu/

- System for annotating, searching and sharing millions of clinical documents, structured data and whole slide images. NLP, code and recode to ontologies, de-identify, etc. to create cohorts.
- TJU just joined the TCRN, StonyBrook soon.
- Regulatory and legal underpinnings lets them operate without and IRB (until tissue is transferred). Conducted through Network Trust Agreements - by definition, not human subjects research.
- Recent Cancer Research publication describing the network
- Focus on API and data analytical framework this year to support broader use of data.
- TCGA demonstration use case (get more information…)

Elmer Bernstam & Funda Meric-Bernstam: Informatics to enable routine personalized cancer therapy https://pct.mdanderson.org/#/home

- Adapt and apply existing informatics techniques and combine with existing “gene sheets”
- Current activities
    - Information curation for decision support
    - AIMED: automated identification of molecular effects of drugs
- Decision support via knowledgebase development and molecular tumor board

- Have created “knowledge sheets” on 27 relevant genes
- Curated over 1500 clinical trials, over 600 drugs
- Important to get the workflow right, ranking is also important
- Not ingesting VCF files; going from the literature

Key Note

Eli Van Allen: Clinical interpretation of the cancer genome for precision oncology

- Genotyping at the point of care to guide individualized patient care
- Precision Heuristics for Interpreting the Alteration Landscape (PHIAL) algorithm. “If this, then that heuristics” aka a decision tree
- Rank alterations by biological and clinical relevance and build a model how to interpret patient’s genome
- Van Allen, Wagle et al. Nature Medicine 2014 (PHIAL publication)
- BRCA2 K3326* - inherited variant for metastatic prostate cancer. Exceptional response w/platinum-based therapy
- PHIAL is open source and lead to many (>20) collaborations. Now part of a CLIA pipeline
- Time to progression is harder to obtain than sequencing the genome (the challenge of mining the EHR)
- Future: Look for actionable cellular states. Clinical-preclinical matchmaker.

IMAT-ITCR Collaboration Opportunities Tony Dickherber: IMAT program

Innovative Molecular Analysis Technologies (IMAT) program

100-120 projects active at any given time, currently 78 R21s, 53 R33s

Grantees have hardware and chemistry expertise: partnership with data management and analysis capabilities of the ITCR investigators?

David Hill

- Use of IMAT to develop ORFeomes, interactomes, edgeotyping
- “Time to think of different isoforms as different genes”
- Development of interactome networks Biophysical interactions >> Functional interactions
- Most of the world is chasing a small corner of the interactome. Need systematic maps. Goals is to create a disease interactome.
- Edgotypes: Not a full knockout, but eliminating some of the interactions. Null vs edgetic vs wild type.
- Disease alleles lead to perturbed interactions

IMAT-ITCR Discussion

- Collaborate on developing QC for new methods and to resolve issues with batch effects. Small supplements to fund these proof of concept projects.
- Address tumor heterogeneity by combining imaging and molecular information.
- Currently have R33 (IMAT) and U24 (ITCR). ITCR work led to the R33. New technique often requires new analytics. Release of these methods should be done through widely used platforms (e.g. Trinity, GenePattern, Galaxy, etc.)

- Longer term: we need to use multiscale methods to drive discovery and treatment decisions. ITCR needs to “Future-proofed” by working with the cutting edge technology development.

Genomics Session III

Rachel Karchin and Mike Ryan: Informatics analysis for high-throughput analysis of cancer mutations CRAVAT http://www.cravat.us/

- Entire CRAVAT system has been packaged into a Docker image (on Docker Hub). includes a user-friendly installation procedure. 376 downloads in just a few months. Also developing new Galaxy tools.
- Funded in 2012. Project goals:
    - Integrate tools to prioritize missense mutations
    - Broaden the tools scientifically
    - 2016: integrated tool with broader scientific scope. Paper published.
- NDEx collaboration: test for enrichment in a library of pathways
- HotMAPS: Clusters of missense mutations that occur together in 3D space. Computed in all TCGA tumor types.
- Have created hooks for integration based on administrative supplements - REST-API
- U24 aims
    - New features
    - interoperability/interactions with other tools including Cloud Pilots and Galaxy tools
    - Keep system up to date. Rebuild for hg38

Bobbie-Jo Webb-Robertson: Interactive Informatics Resource for Research Driven Cancer Proteomics

- Goal: support anyone who wants to explore peptide/protein interactions in cancer
- Isoforms big challenge in proteomics
- Need to deal with the missing data problem - combination of technology plus limit of detection issues. Imputation is a problem. Need methods that can operate in the context of missing data (e.g. PCA).
- P-Mart - make proteomics stats methods available in R. Using protein and gene data as provided by CPTAC DCC.
- http://pmart.azurewebsites.net/pmart/index.html
- Collaborating with various CPTAC centers. Developing interface to upload researchers own data. Uses R data frame. Essentially two spreadsheets that are needed.
- Trelliscope

Yaoyu Wang: MeV: Software for Next Generation Data Analysis http://mev.tm4.org/#/welcome

- Since 2007 more than 204K downloads from SourceForge. Now moving to the cloud to address the many problems/challenges associated with standalone applications
- Implementing Rserve client on AWS compute node
- Dependency injection with Raven

- OpenRefine for cohort selection
- Integrating with the Cloud Pilots to streamline TCGA data access; also with VisANT and Cytoscape for network visualization and analysis
- Audience question: Do you support workflows? This would be of interest.

Jingshan Huang: OmniSearch

- GitHub: https://github.com/OmniSearch
- Web search interface: http://omnisearch.soc.southalabama.edu/ui/
- Wiki: http://omnisearch.soc.southalabama.edu/w/index.php/Main_Page
- Helps to identify role of microRNAs in various human cancers
- Created miRNA domain ontology - OMIT. Part of OBO Foundry and NCBO Bioportal
- Collaborating widely with the bio-ontology community including GO, SO, PRO, CHEBI, OBI, and UBERON

Imaging Session II

Gordon Harris: web viewer for oncologic imaging research

- Tumor Imaging Metrics Core, approved as an NCI Core Resource in 2006
- Over 700 active trials
- Geared for the site level read for cancer centers
- http://www.precisionmetrics.org. Implemented at 5 other cancer centers
- Metrics Manager: Built on ClearCanvas. Challenging to deploy, only compatible with Windows, must be on hospital network >> Open Health Imaging Foundation. 501(c) 3. Ohif.org. Building LesionTracker under this. http://leasiontracker.ohif.org
- Goals: Zero-footprint, secure, state-of-the-art, open source
- Built on Cornerstone, dicomParser
- Pulling some functionality out of ClinicalMeteor project
- Integrator with 3D Slicer and many other potential ITCR/QIN collaborations
- Browser question: Do have to be IE compatible. Aiming for IE11, hoping not to have to support IE9.

Joel Saltz: Tools to analyze morphology and spatially mapped molecular data

- Suite of five tools to integrate radiology, pathology and molecular data.
- Pathomics/Histomics: ID and segment trillions of objects, feature extraction, analysis
- Aims: Pipelines, Database, HPC software, visualization middleware
- Focus on QC: Developed tools to curate data, heatmap to visualize agreement between algorithms.
- FeatureScape, FeatureExplorer
- 3D slicer pathology - Generate high quality ground truth
- http://quip1.bmi.stonybrook.edu

Ken Huang: Informatics links between histological features and genetics in cancer

- NCI CPTAC contract (2 year) compliments the ITCR grant. Proteomics data.integration.
- Focus on integrating genomics, clinical, proteomics and histopathology data

- http://osumo.org - initial application
- Can work locally until more computing is needed, then work remotely
- Input: Any Girder-uploaded dataset
- iGPSe: Visual analytic system for integrative analysis
- https://github.com/osumo/osumo
- Academic-industry partnership: Kitware's focus is on software engineering, OSU focus is on the science

Anant Madabhushi: Pathology image informatics platform for visualization, analysis and management

- PIIP is based on the Sadeen imagine viewer; plugin architecture
- Can execute MatLab routines from within the viewer. Expect much of the code that will come in will be MatLab
- Goal is to improve the Sadeen API - improved plugin framework (among others). Provide an app store where the community begins to contribute. Analogy to 3D Slicer.
- Also an academic-industry partnership (with Pathcore)
- Color normalization is the first application developed - critical need
- Biomarker quantification algorithm embedded in the Sadeen viewer
- Nuclei segmentation: Using a method developed by Metin Gurcan
- Co-registration of IHC slides, tools to compare multiple IHC stains.
- Human factors engineering to improve usability
- 1st release of updated Sadeen viewer released 2/23/16; next anticipated in June, 2016
- Goals: Tissue Microarray spot extraction, add nuclei segmentation algorithms

Technical WG Working Lunch

- General interest in merging to a single Techical WG
- What are the Scientific motivations for new integration projects?
- Work on sharing and finding containers more easily
- We need Docker custom labels than can be parsed to understand information. Self-describing formats. DockStore is doing this and has APIs for searchability.
- Allows you to use the container without having to follow up with the developer.
- Are there any other integrative approaches besides Docker?
- Metadata can be independent of the use of Docker.
- Is there some potential for engaging with data generating studies to drive some ITCR efforts? Framework for plug and play.
- TCGA data has its limitations. Have had to go to cooperative groups to get access to datasets. Process to access data took 2 years.
    - Formal proposal that is reviewed within NCI is very laborious. This is an opportunity to engage. Good for predictive studies.
- Are there nearer-term approaches? Are there 1 or 2 datasets we can get access to? E.g. UPitt breast cancer data that has outcomes data and treatment data.
- Make the case that we need the data to advance the science.
- EDRN NLST
- **ECOG/ACRIN: since they are part of QIN.

- Lymphoma Consortium data.
- Need genomics, proteomics, outcomes, histopathology, radiology, immunohistochemistry data
- Take advantage of the work Rebecca has done in terms of data sharing.
- >>Send Rebecca all of the ITCR sites
- Marker of benefits for a drug treatment would be a good use case, for example.
- How can this group influence the challenges with data sharing. Reminder to submit ideas to the Moonshot through http://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/get-involved

Genomics Session IV

Benjamin Berman: Software tools for regulatory analysis of large cancer methylome datasets

- Tools for genome scale methylation analysis
- ELMER http://bioconductor.org/packages/devel/bioc/html/ELMER.html
- Transcription factors are ~25% of significantly mutated genes in TCGA
- TF binding site analysis of mutations in cancer (cis-acting) - motifbreakR
- StateHub: std exchange format for combinatorial chromatin states. Fully versioned. AVailable on GitHub https://github.com/Junkdnalab/StateHub and /StatePaintR
- Next steps: tool integration to investigate trans and cis mutations affect on TF networks, StateHub.
- BioC, GitHub, and Docker links available at junkdnalab.org

Maureen Sartor: mint and annotatr: assessing DNA methylation and hydroxymethylation signatures in cancers

- Mint - methylation integration. Command line as well as Galaxy implementation
    - Now has a rule-based classification scheme for classifying CpG sites
- annotatr - R package for annotation and viz
    - Annotation of genomic regions. Simple, fast, flexible
- PePr - differential binding
- methylSig - differential methylation for bisulfide sequencing data
- 5hmC - correlates better with gene expression than 5mC
- Assay methods exist to identify one, the other, and both
- New features include common SNP filter
- Galaxy tools integrated for mint, Galaxy workflows
- Integration with RNAseq data.

John Weinstein and Bradley Broom: Next-generation clustered heatmaps for fluent, interactive exploration of omic data

- Visualization tool for large omics datasets, generate publication quality graphics, dynamic link-outs to other resources (NCBI, Genecards, cBio portal, cancer digital slide archive), dynamic analyses sites (Box plot, Points plot).
- Docker-compose NG-CHM container system simplifies deployment. Also for RStudio container with NG-CHM
- New graphical builder for the NG-CHM for ISB-CGC

- New architecture to take advantage of new browser capabilities. Allows the CHMs to be saved locally. Added to Galaxy for BigQuery interface to ISB-CGC
- Next year: new architecture NG-CHMs, scale up to larger data matrices, tablet optimized user interface
- Any clustering you can do in R, you can represent in the NG-CHM. Also, doesn't have to be clustering

Big Data and EMR Session

Joshua Stuart, Kyle Ellrott: ESCE: Discovering molecular processes

- Biomedical Evidence Graph (BMEG)
- Organize crowdsourced challenges: old and new DREAM challenges
    - Old methods: submit predictions (user runs the method)
    - New methods: submit predictors (judge runs the method)
- Bake-off analogy: you don't just get the cookies, you also get the ingredients and the recipes
- DREAM inspires compatibility - need common inputs and outputs
- MC3 - TCGA somatic mutation calling. Open MAF posted last Friday, 6/10/16
- RABA code: Runnable by anyone anywhere. To get RABA code, changed the way they do DREAM challenges.
- Partnering with the NCI cloud pilots for an SMC RNA-Seq challenge
- California Kids Cancer Comparison

Hong Yu: EMR adverse drug event detection for pharmacovigilance

- Deep learning for medical event detection in EHRs
- Challenges: lno common vocabulary between different medical entities, unbalanced data, label sparsity. Noisy incomplete sentence structure.
- Recurrent Neural Networks (RNN) remembers previous context history, do not need handcrafted features. Able to recognize patterns in variable context lengths.
- Bidirectional RNNs
- New work: Improve scores, map entity to controlled terminology

Guergana Savova, Rebecca Crowley Jacobson: Cancer deep phenotype extraction from electronic medical records (DeepPhe)

- Extract all patient phenotype information from EMR
- http://cancer.healthnlp.org
- Using cTAKES as the NLP annotator for the document pipeline
- Storing results in neo4j (to drive visualization) and TransMart
- FHIR information model allows interoperability with many systems
- Agile software development
- Grant supplement to work with NCI SEER to extend DeepPhe methods for cancer surveillance
- Goals for next year are temporal relations, expansion to ovarian cancer, include more visualizations.

Imaging Session III

Hugo Aerts: Quantitative radiomics system: decoding the tumor phenotype

- Tumors are different and medical imaging can capture these phenotypic differences
- Integration of molecular imaging, functional imaging, anatomical imaging
- Workflow: Imaging, segmentation, feature extraction, and analysis
- Radiomics feature set includes stats, texture, shape (current release ~1600 features)
- Radiomics-genomics association modules:13 association modules identified and validated in cited study
- Demonstrated that radiomics strengthens molecular signatures
- Image quality has to be optimized; accurate segmentation, generalized technical platforms needed, feature definitions have to be generalized.
- Radiomics imaging platform implemented in 3D Slicer
- PyRadiomics - runs standalone and within 3D Slicer (release fall 2016)
- Horizon - server for medical image processing. Built on Cornerstone (OHIF) and GIRDER

Lauren O'Donnell: Open source diffusion MRI technology for brain cancer research

- Unique non-invasive MRI, measures 3D profile of water diffusion. Faster diffusion along axons.
- Tractography: measurement of the brain's white matter tracts
- Enhance 3D Slicer SlicerDMRI
- Creating the first DICOM standard for tractography. Soon to be available through DCMTK: open source DICOM library
- Many challenges, one of which is identification of important tracts. Working on automated tract identification - multi-fiber tractography

Discussion of Collaboration and Integration Opportunities

- Requests: Post participant list, presentations and posters on the NCIP Hub.
- Integration and interoperability of genomic analysis tools, integration of genomic and image analysis tools.
- Need to aggregate tools within domains.
- Specific challenge: people choose their partners.
- Need to take a look at the latest DREAM challenge: co-operative model may work better for the ITCR rather than a challenge.
- Follow-up: Define and identify attributes for a dataset to collaborate on.